REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)* 31-05-2013	2. REPORT TYPE Journal Article	3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE Spatial Error Metrics for Oceanographic Model Verification	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 0602435N
6. AUTHOR(S) Sean B. Ziegeler, James D. Dykes, and Jay F. Shriver	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER 73-4320-01-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory Oceanography Division Stennis Space Center, MS 39529-5004	8. PERFORMING ORGANIZATION REPORT NUMBER NRL/JA/7320-11-738

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research One Liberty Center 875 North Randolph Street, Suite 1425 Arlington, VA 22203-1995	10. SPONSOR/MONITOR'S ACRONYM(S) ONR
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A common problem with modern numerical oceanographic models is spatial displacement, including misplacement and misshapenness of ocean circulation features. Traditional error metrics, such as least squares methods, are ineffective in many such cases; for example, only small errors in the location of a frontal pattern are translated to large differences in least squares of intensities. Such problems are common in meteorological forecast verification as well, so the application of spatial error metrics have been a recently popular topic there. Spatial error metrics separate model error into a displacement component and an intensity component, providing a more reliable assessment of model biases and a more descriptive portrayal of numerical model prediction skill. The application of spatial error metrics to oceanographic models has been sparse, and further advances for both meteorology and oceanography exist in the medical imaging field. These advances are presented, along with modifications necessary for oceanographic model output. Standard methods and options for those methods in the literature are explored, and where the best arrangements of options are unclear, comparison studies are conducted. These trials require the reproduction of synthetic displacements in conjunction with synthetic intensity perturbations across 480 Navy Coastal Ocean Model (NCOM) temperature fields from various regions of the globe throughout 2009. Study results revealed the success of certain approaches novel to both meteorology and oceanography, including B-spline transforms and mutual information. That, combined with other common methods, such as quasi-Newton optimization and land masking, could best recover the synthetic displacements under various synthetic intensity changes.

15. SUBJECT TERMS

Error analysis, Forecast verification, Ocean models

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON James D. Dykes
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	07	19b. TELEPHONE NUMBER *(Include area code)* (228) 688-5787
Unclassified	Unclassified	Unclassified			

Spatial Error Metrics for Oceanographic Model Verification

SEAN B. ZIEGELER

High Performance Technologies, Inc., Reston, Virginia

JAMES D. DYKES AND JAY F. SHRIVER

Naval Research Laboratory, Stennis Space Center, Mississippi

ABSTRACT

A common problem with modern numerical oceanographic models is spatial displacement, including misplacement and misshapenness of ocean circulation features. Traditional error metrics, such as least squares methods, are ineffective in many such cases; for example, only small errors in the location of a frontal pattern are translated to large differences in least squares of intensities. Such problems are common in meteorological forecast verification as well, so the application of spatial error metrics have been a recently popular topic there. Spatial error metrics separate model error into a displacement component and an intensity component, providing a more reliable assessment of model biases and a more descriptive portrayal of numerical model prediction skill. The application of spatial error metrics to oceanographic models has been sparse, and further advances for both meteorology and oceanography exist in the medical imaging field. These advances are presented, along with modifications necessary for oceanographic model output. Standard methods and options for those methods in the literature are explored, and where the best arrangements of options are unclear, comparison studies are conducted. These trials require the reproduction of synthetic displacements in conjunction with synthetic intensity perturbations across 480 Navy Coastal Ocean Model (NCOM) temperature fields from various regions of the globe throughout 2009. Study results revealed the success of certain approaches novel to both meteorology and oceanography, including B-spline transforms and mutual information. That, combined with other common methods, such as quasi-Newton optimization and land masking, could best recover the synthetic displacements under various synthetic intensity changes.

1. Introduction

The concept of spatial error was introduced by Hoffman et al. (1995), who proposed two general types of error, including spatial (or displacement) and intensity (or amplitude) errors. Spatial error metrics have since advanced in a number of studies (Casati et al. 2008; Gilleland et al. 2009, 2010a; Ahijevych et al. 2009; Marzban et al. 2009). More recent developments include an alignment method by Beechler et al. (2010) only operating along one dimension. Clark et al. (2010) searches for the best matching point within a predefined window size. Marzban and Sandgathe (2010) utilize a variation of optical flow to improve the handling of intensity differences. Gilleland et al. (2010b) use thin-plate splines to deform one dataset to match the other.

These methods first determine a spatial displacement to serve as one error metric. Many go further by correcting the displacement and then consider any remaining difference an intensity error. This separation of errors can provide key information for numerical prediction model validation. The intensity error can be displayed qualitatively by simply displaying the point-wise difference magnitude as an image. Gilleland et al. (2010a) illustrates these difference images before and after correcting displacement to show how true intensity errors are discovered.

The above methods were applied to meteorological datasets. The application of displacement metrics to oceanographic data seems sparse, yet we assert that it would be beneficial. The use of traditional statistical error metrics for oceanographic modeling is common, including root-mean square (RMS), cross correlation (CC), mean bias,

Corresponding author address: Sean Ziegeler, High Performance Technologies, Inc., 11955 Freedom Dr., Suite 1100, Reston, VA 20190-5673.
E-mail: sziegeler@hpti.com

and skill scores (Murphy and Epstein 1989; Murphy 1995; Helber et al. 2010). Mariano (1990) tracks the motion of isocontours as displacement error. Hurlburt et al. (2008) manually identify features in satellite imagery and model results and compare the positions by hand. A method similar to Clark et al. (2010) is utilized by Helber et al. (2010) such that data points are grouped into analysis windows. This paper presents the hypothesis that algorithmically detected displacement for ocean forecast results can be more accurate than traditional statistical metrics, more comprehensive and less laborious than manual feature tracking, and provides both quantitative and qualitative results.

Both meteorology and oceanography could benefit from the advanced image alignment methodologies, referred to as "registration," used in the medical imaging field. Registration can occur, for example, between computed tomography (CT) scans from different times, different patients, or to images of an entirely different modality such as magnetic resonance imaging (MRI). The field was already under heavy development by 1998 as exemplified in the survey by Maintz and Viergever (1998). This paper focuses on "deformable" registration, where the alignment can be nonuniform and tends to vary significantly. Overall, such methods can be divided into nonparameterized methods (e.g., optical flow) and parameterized, which use a parametric function (e.g., splines) to model the displacement. Registration in both medical imaging data (Crum et al. 2004) and meteorological models (Ahijevych et al. 2009) favors "multiscale" approaches, where the two datasets are successively processed from a coarse resolution to the finest. One can also utilize masks to exclude invalid or irrelevant data. A registration system is divided into three components: 1) transform, 2) difference criterion, and 3) optimizer. The transform deforms a trial dataset to appear like the reference dataset. The difference criterion (DC) measures the difference between the datasets. The optimizer determines how to adjust the transform to optimize the DC. The three components iterate until the optimizer decides convergence is achieved (Fig. 1). Meteorological applications for displacement error have almost exclusively used a form of RMS or absolute error as the DC. Medical image registration introduces normalized correlation (NC), mutual information (MI), and others.

2. Registration method

A parameterized approach was chosen since nonparameterized methods (e.g., optical flow) have not enjoyed much success with data of significantly varying intensities (Crum et al. 2004). Of the parameterized approaches, the B-spline transform seems the most common (Crum et al.
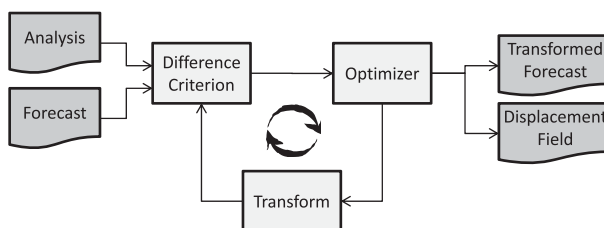

FIG. 1. The three components of registration, difference criterion, optimizer, and transform, looping until completion. The inputs are the analysis and forecast datasets. The outputs are a transformed (corrected) forecast to best match the analysis and the displacement field used to apply the transform.

2004). A B-spline transform is governed by control points connected by 2D B-spline curves (Fig. 2a), and can be evaluated to displacement vectors at every data point (Fig. 2b). B-spline curves offer two advantages over the thin-plate splines from Gilleland et al. (2010b). First, B-splines have implicit smoothness constraints guaranteed by their continuity and differentiability properties. Thin-plate splines require a penalty parameter to reduce sharp edges in the splines, and the correct value of the parameter can require some effort to determine initially for different applications. B-splines also have a local region of influence, allowing DC calculations to be calculated over smaller regions for efficiency (Crum et al. 2004). Multiscale processing is provided as an option to facilitate the detection of larger spatial errors by starting the registration with a coarse sampling (reduced recursively by some power of two) of the original grids. When the registration has converged at this coarse scale, it is resumed on a finer grid (finer by a factor of 2). This process is repeated until the registration has been performed at the original resolution. A related option includes scaling the control point spacing at each of the aforementioned grid scales. Finally, the masking option is implemented geometrically across all scales to allow the treatment of only the valid points.

A number of optimizers exist (Crum et al. 2004), but it is important to note that every B-spline point adds more degrees of freedom. The gradient descent optimizer tends to converge slowly for more than a few degrees of freedom. Stochastic and evolutionary optimizers are good for locating global optima but can be slow to converge. Conjugate-gradient and quasi-Newton methods are intended for many degrees of freedom and are simple to initialize.

The most common DC is mean-square error MSE (the square root of MSE, RMSE, adds extra calculations unnecessary for optimization), which follows
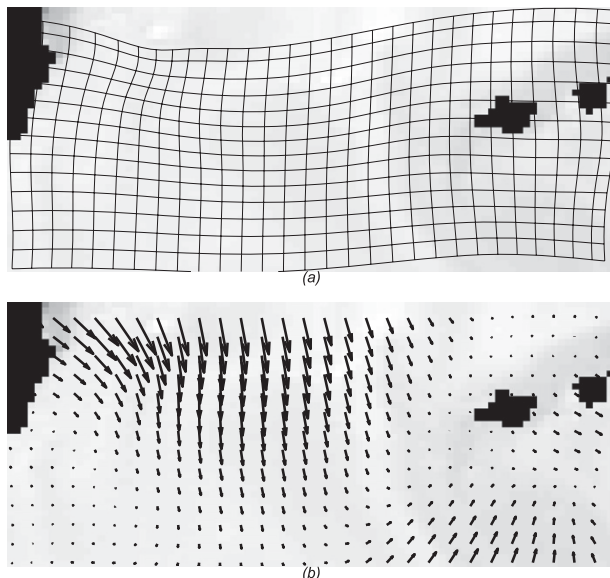
$$\text{MSE} = \frac{1}{n}\sum_{x,y}[f(x,y) - g(x,y)]^2,$$

FIG. 2. (a) An example of a B-spline transform. (b) The same B-spline transform converted to a displacement vector field.

where x,y span the dimensions, and $f(x, y)$ and $g(x, y)$ are the reference and trial datasets, respectively. The NC is also common:

$$\text{NC} = \frac{1}{n-1} \sum_{x,y} \frac{[f(x,y) - \overline{f}][g(x,y) - \overline{g}]}{\sigma_f \sigma_g},$$

where $\overline{f}$ and $\overline{g}$ are the respective dataset means, and σ_f and σ_g are the respective deviations. MSE can be very susceptible to intensity differences. NC improves this, but typically only for linear intensity differences (Pluim et al. 2003).

Viola and Wells (1995) introduced MI as a DC for "multimodal" registration—that is, registration across different image modalities such as CT and MRI. MI is defined as follows:

$$\text{MI} = H[f(x,y)] + H[g(x,y)] - H[f(x,y), g(x,y)],$$

where $H(f)$ is the marginal entropy of random variable f and $H(f, g)$ is the joint entropy. In this case, the random variables are the datasets themselves, either sampled at every data point or a representative random sample of the data points. The marginal and joint entropies are defined as follows:

$$H[f(x,y)] = -\int_{x,y} p(f) \log p(f)\, df, \quad \text{and}$$

$$H[f(x,y), g(x,y)] = \int\int_{x,y} p(f,g) \log p(f,g)\, df\, dg,$$

where $H(f)$ is also applied for $H(g)$, $p(f, g)$ is the joint probability density function (PDF) and $p(f)$ and $p(g)$ are the respective marginal PDFs. The marginal PDFs can be computed approximately using a histogram, where each histogram bin count is divided by the total number of samples to form a bin probability. This reduces the marginal entropy calculations $H(f)$ and $H(g)$ to merely the sum of the bin probabilities multiplied by their respective logarithms. The joint PDF is estimated by a two-dimensional histogram, where each bin represents a pairing of two values, each respectively from the same location in the datasets f and g. As with marginal entropy, the joint entropy is the sum of all bin probabilities multiplied by their respective logarithms.

Compared to MSE and NC, MI should handle nonlinear intensity differences better because of the nonlinear nature of the PDFs in relation to each other. Other research has improved MI sophistication as documented in Pluim et al. (2003). A combination of improved MI with B-spline transforms is presented in Mattes et al. (2003). MI was improved by using kernel density estimators (also known as Parzen windows) as a more advanced form of histogram. This was successful for multimodal registration, suggesting that MI would be superior in comparing model results to other sources of reference data such as satellite imagery.

The system presented here is designed to compare a model forecast field to an analysis field resulting from data assimilation. Hoffman et al. (1995) used this approach, and more recent examples include Casati et al. (2008) in meteorology and Wallcraft et al. (2002) in oceanography. The advantages of using an assimilative analysis field from the same model [rather than actual ground truth (e.g., satellite imagery)] include matching grids and similar scalar properties. The disadvantage is that it will miss errors in the model's assimilation process itself. To test the system, in the trials in sections 4 and 5, the reference data is an analysis and the trial forecast is synthetically deformed from that analysis. The registration result is compared to the synthetic deformation for accuracy.

3. Parametric arrangements

Pretrials for a small number of datasets (section 4) eliminated obviously poor options and arrangements of those options. Final trials were run on more datasets (section 5) to find an ideal arrangement. Both trials utilize the Insight Segmentation and Registration Toolkit (ITK; Yoo et al. 2002; Yoo 2004), which provides several DCs and optimizers.

Ng (2005) provides a heuristic starting point for choosing arrangements. Control point spacing for the B-spline curves should not be too coarse or fine. Multiscale is

usually necessary; moreover, the number of scale levels should be such that the lowest grid scale's size is roughly 64×64—that is, 64 grid points in each direction. One should also consider enabling the option to scale the control point spacing with each grid scale. It is expected that the quasi-Newton or conjugate-gradient optimizers would be the best, but optimization can vary depending upon the problem set, so several options were tested including a stochastic optimizer and an evolutionary optimizer. MI was expected to be the best DC, but mean square (MS) and NC were tried also. In addition, a smoothed-gradient version of each MS, NC, and MI are tried as a DC. As the registration progresses, the DC must also interpolate from the forecast data's transformed points to the original gridpoint locations of the analysis field. Either linear or cubic interpolations are appropriate, and one must balance the faster evaluations of the former against the typically faster convergence of the latter. Finally, the use of a land mask is expected to improve registration, but seems undocumented with regards to oceanographic data, so it should be optional in the trials.

The synthetic forecasts are created by applying a deformation field to the analysis, where the deformation field is the Gaussian-smoothed ocean current field from the corresponding analysis. The vector field is scaled in magnitude to be 11.25 times the grid spacing ($\frac{1}{8}°$) per 1 m s^{-1}. This parallels one method for verification with synthetic deformations in medical image registration using historically documented physical motion transforms (Crum et al. 2004). Here, the smoothed current field is used as an approximation of the advection of an ocean model over several time steps. The performance of any given registration arrangement is judged by the nearness of the displacement field recovered by the registration to the original synthetic deformation field. Nearness is defined as the smaller RMS of the vector subtraction between the vector fields. The performance of a registration arrangement under intensity changes is also evaluated by adding functions in conjunction with the synthetic deformation. In this study there are four added functions: zero, constant, low-frequency sinusoid, and high-frequency sinusoid. The maxima of the functions added are one-tenth of a standard deviation of the surface temperatures of the respective dataset, which therefore serves as the single value for the constant functions and as the maximum amplitudes of the sinusoids.

4. Pretrial results and adaptations

The pretrials provided the intended information, which included 1) the elimination of some arrangement options, 2) highlighting which options would require full trials to eliminate or include, and 3) discovery of shortfalls that

would require modifications before proceeding to full trials. First, B-spline control point spacing was too fine at 4 control points per data point, and too coarse at 10 control points per data point. Both 6 and 8 points were better, but it was unclear which of the two was best, leaving that determination for the full trials. Multiscale processing was indeed superior, with an average of 64×64 grid points being generally the best size for the initial coarsely scaled grid. Since most grids are not powers of 2 in size, and each scale is half the size of the finer scale, a range of the coarsest scale is expected. We found that less than 45×45 was too small and over 90×90 was too large. Results were also consistently better when enabling the option to scale control point spacing along with the gridpoint scale. The choice between linear and cubic interpolation for the DC was inconclusive.

The best optimization strategy was quasi-Newton. The quality of the end results were all relatively the same, but the speed of convergence was the deciding factor. Quasi-Newton was also both easier to configure and tended to adapt to different datasets quickly. Conjugate gradient was a close second overall, with speeds of convergence as much as 20% slower. Methods without DC derivatives such as stochastic perturbation and an evolutionary optimizer were much too slow to converge—slower than quasi-Newton by orders of magnitude. Quasi-Newton also allows freezing control points over land areas, but this tended to restrict the movement of the B-spline grid in water areas near land and prevented optimal solutions in those areas, so this option was excluded.

MI seemed like the best DC, but the result was not conclusive nor was using the gradient in conjunction with the DCs. It was determined that, for MI, the histogram size must be large enough—that is, somewhere near 64 bins at the minimum scale of a multiscale run. Also, it must resize in proportion with the largest dimension of each scale of a multiscale run. Thus, since each scale doubles in size along both dimensions from the coarsest scale, the histogram size must also double at each scale.

Finally, land masks were indeed useful but only when the masked areas are handled properly. Masked values must be excluded from all DC aspects, including minimum/maximum calculations, and left out of interpolations used to generate the multiscale levels. It was also necessary to overwrite masked land areas with a first-differentiable boundary condition to reduce errors in the B-spline evaluations near land.

5. Full trial results

The pretrial results left the following arrangement parameters undetermined: 1) six versus eight data gridpoint spacing per B-spline control point, 2) linear versus cubic
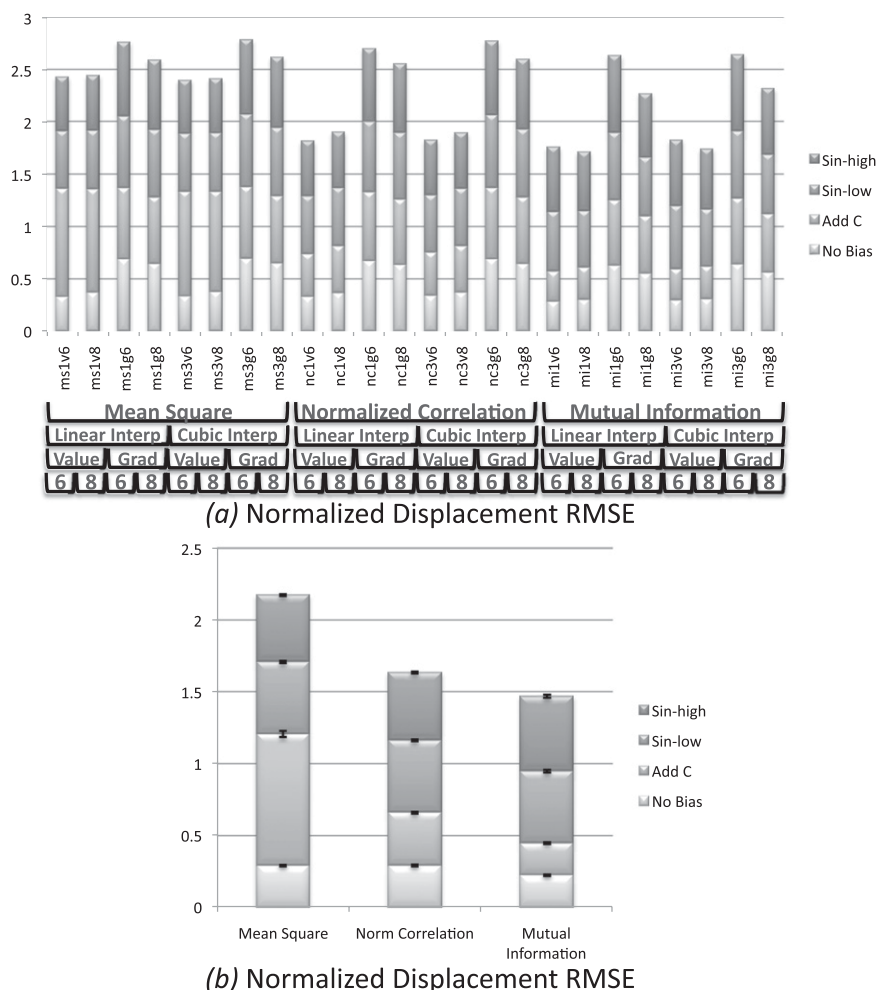
(a) Normalized Displacement RMSE

(b) Normalized Displacement RMSE

FIG. 3. (a) All results from the full trials. The codes along the bottom indicate the registration options for the respective arrangement (i.e., "ms" is mean square, "nc" is normalized correlation, "mi" is mutual information, "1" is linear, "3" is cubic, "v" is direct value, "g" is gradient, "6" is six control points per data point, and "8" is eight control points). The brackets underneath help visually categorize those arrangements. The stacked items indicate the added synthetic intensity function. (b) The same results data condensed into the three displacement criteria by keeping only the best optimized criterion for each.

interpolation, 3) MSE versus NC versus MI for DC, and 4) direct DC versus smooth-gradient DC. The full trial ran every combination of the above options on each dataset with the four intensity change functions. The study used 20 datasets from 24 times (2 per month) throughout 2009. Each dataset is a cutout of the global $\frac{1}{8}°$ Navy Coastal Ocean Model (NCOM) surface temperature field. The model itself is on an orthogonal curvilinear horizontal grid. Each cutout is interpolated onto a rectilinear grid of $\frac{1}{8}°$ resolution and rectangular shape from various locations around the world and varies in size, scalar range, and features. The grid sizes range from 97 × 41 to 481 × 561 with a mean size of 421 × 290. Results are in relative RMS of vector difference between synthetic deformation

and registered displacement (i.e., RMS normalized by the RMS of deformation magnitudes for a given dataset).

The accumulation of results is in Fig. 3a. The smooth-gradient-based DCs are clearly inferior. To simplify further interpretation, each result that best optimizes its respective DC can represent that DC for each arrangement and dataset, resulting in Fig. 3b. Overall, MI has the best accuracy, but some datasets did show decreased accuracy nonetheless. These were regions with both significant temperature gradients and large areas of near-constant temperature (e.g., combined equatorial and polar areas).

Qualitative results demonstrate the improvement realized when correcting for displacement. Figure 4a shows
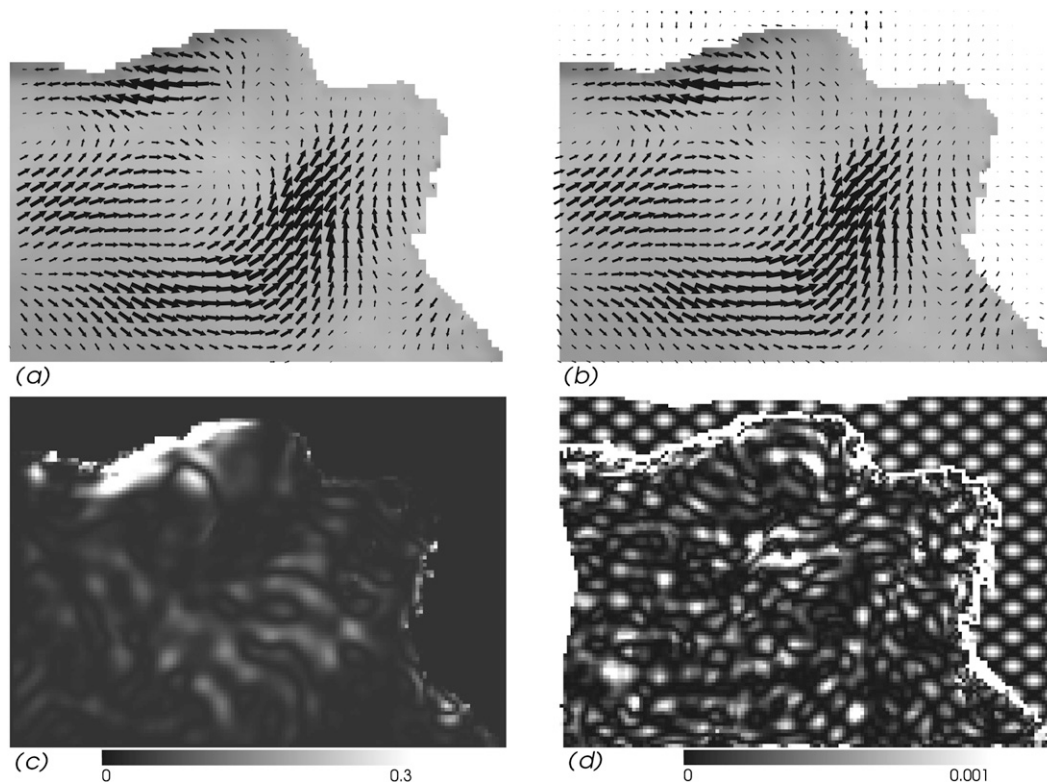
FIG. 4. (a) The original analysis dataset and the synthetic deformation displayed as vectors. (b) The synthetic forecast created by the deformation and the displacement field as recovered by the registration. (c) The absolute point-wise difference between the analysis and synthetic forecast. (d) The absolute point-wise difference between the analysis and displacement-corrected forecast. The emerging pattern is the synthetic high-frequency sinusoid applied to test the registration process under intensity changes.

the synthetic deformation, and Fig. 4b shows the registration displacement. Figure 4c is the difference magnitude between the analysis and synthetic forecast. The high-gradient region near the northwest shore dominates the error. This high-gradient region was (synthetically) predicted, but the location was shifted by the displacement. Correcting for the displacement in Fig. 4b then repeating the absolute difference produces what is depicted in Fig. 4d. Notice the significantly smaller absolute difference values on the color scale from Figs. 4c to 4d indicating an improvement in overall error measurement. It is also now obvious that this particular dataset had its intensity synthetically altered by the high-frequency sinusoid as it is now mostly recovered in Fig. 4d.

6. Discussion and future work

MI without a smooth-gradient modification was the best overall DC. As theoretically expected, MI handled low local entropy (e.g., low-frequency sinusoid) intensity changes better than high local entropy changes, but it functioned at least as well as other DCs under high local entropy changes such as the high-frequency sinusoid addition. One improvement for MI would be adaptive kernel density estimation, which would allow for variable-sized histogram bins. This would likely solve the accuracy problem in combined equatorial and polar regions as long as the variable bins' sizes were kept the same between the two compared datasets.

The best choice for interpolation order and B-spline grid spacing varies given the dataset, but one has the option of running them all and choosing the best optimized DC. Finally, proper treatment of land masks is critical. The results show that the registration process, once tuned, works quantitatively and qualitatively for this oceanographic data and successfully separates the model error into displacement and intensity components.

This study was limited to $\frac{1}{8}°$ cutouts of surface temperature from a global model, and it would be difficult to predict its applicability to other situations without the references cited in this paper. As such, the prior art in the field of meteorology suggests that spatial error via registration in general would be useful for a variety of resolutions and variables. It should even be possible to apply it

to global model results, though the aforementioned problem with datasets containing both equatorial and polar regions must be addressed first. The prior art in medical imaging shows even more promise, comparing multimodal datasets, including those with widely varying resolutions and scalar properties. This suggests that registration could be easily applied to more than just temperature for spatial error. It should even be possible to apply registration to several variable fields at once using multivariate mutual information, but this would be a topic for future work. Registration has been applied to three-dimensional (3D) medical imagery, and should be possible for oceanographic models as well, though one would likely need to make special modifications to handle the often-used nonuniform spacing between depth layers, as that is not an issue in medical imaging. Finally, while this study was intended to improve the analysis of forecast skill, it should also be applicable for simulation errors in data assimilation systems, though one would need to determine how to incorporate the spatial errors to improve the simulation errors. For example, it would be possible to apply the inverse of the displacement field to "correct" future assimilation results.

This study aims to serve as a first step in establishing how spatial error might be best applied in oceanographic modeling. It is important to note that the synthetic errors generated in the study may be significantly different to actual model biases in various scenarios. Thus, follow-up work for the authors includes performing a study comparing multiple oceanographic models to actual ground truth data (e.g., satellite imagery). The study will run alongside the same comparisons using traditional error metrics to determine which error metrics provide the best portrayal of forecast skill.

REFERENCES

Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. *Wea. Forecasting,* **24,** 1485–1497.

Beechler, B. E., J. B. Weiss, G. S. Duane, and J. Tribbia, 2010: Jet alignment in a two-layer quasigeostrophic channel using one-dimensional grid warping. *J. Atmos. Sci.,* **67,** 2296–2306.

Casati, B., and Coauthors, 2008: Forecast verification: Current status and future directions. *Meteor. Appl.,* **15,** 3–18.

Clark, A. J., W. A. Gallus Jr., and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Wea. Forecasting,* **25,** 1495–1509.

Crum, W. R., T. Hartkens, and D. L. G. Hill, 2004: Non-rigid image registration: Theory and practice. *Br. J. Radiol.,* **77,** S140–S153.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting,* **24,** 1416–1430.

——, ——, ——, and E. E. Ebert, 2010a: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.,* **91,** 1365–1373.

——, J. Lindström, and F. Lindgren, 2010b: Analyzing the image warp forecast verification method on precipitation fields from the ICP. *Wea. Forecasting,* **25,** 1249–1262.

Helber, R. W., J. F. Shriver, C. N. Barron, and O. M. Smedstad, 2010: Evaluating the impact of the number of satellite altimeters used in an assimilative ocean prediction system. *J. Atmos. Oceanic Technol.,* **27,** 528–546.

Hoffman, R. N., Z. Liu, J. Louis, and C. Grassoti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.,* **123,** 2758–2770.

Hurlburt, H., and Coauthors, 2008: Eddy-resolving global ocean prediction. *Ocean Modeling in an Eddying Regime, Geophys. Monogr.,* Vol. 177, Amer. Geophys. Union, 353–382.

Maintz, J., and M. A. Viergever, 1998: A survey of medical image registration. *Med. Image Anal.,* **2,** 1–36.

Mariano, A. J., 1990: Contour analysis: A new approach for melding geophysical fields. *J. Atmos. Oceanic Technol.,* **7,** 285–295.

Marzban, C., and S. Sandgathe, 2010: Optical flow for verification. *Wea. Forecasting,* **25,** 1479–1494.

——, ——, H. Lyons, and N. Lederer, 2009: Three spatial verification techniques: Cluster analysis, variogram, and optical flow. *Wea. Forecasting,* **24,** 1457–1471.

Mattes, D., D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, 2003: PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging,* **22,** 120–128.

Murphy, A. H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting,* **10,** 681–688.

——, and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.,* **117,** 572–582.

Ng, L., 2005: Overview: ITK registration methods. Preprints, *Medical Image Segmentation and Registration with ITK,* San Diego, CA, SPIE, 85 pp.

Pluim, J., J. Maintz, and M. Viergever, 2003: Mutual-information-based registration of medical images: A survey. *IEEE Trans. Med. Imaging,* **22,** 986–1004.

Viola, P., and W. Wells, 1995: Alignment by maximization of mutual information. *Proc. Fifth Int. Conf. on Computer Vision,* Los Alamitos, CA, IEEE Computer Society, 16–23.

Wallcraft, A., H. Hurlburt, E. Metzger, R. Rhodes, J. Shriver, and O. Smedstad, 2002: Real-time ocean modeling systems. *Comput. Sci. Eng.,* **4,** 50–57.

Yoo, T. S., Ed., 2004: *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis.* A K Peters, 426 pp.

——, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, 2002: Engineering and algorithm design for an image processing API: A technical report on ITK—The Insight Toolkit. *Stud. Health Technol. Inf.,* **85,** 586–592.